A Cyber-War Between Bots: Human-Like Attackers are More Challenging for Defenders than Deterministic Attackers

Yinuo Du
Carnegie Mellon University
yinuod@andrew.cmu.edu

Baptiste Prebot
Carnegie Mellon University
bprebot@andrew.cmu.edu

Xiaoli Xi
Carnegie Mellon University
xiaolix@andrew.cmu.edu

Cleotilde Gonzalez
Carnegie Mellon University
coty@cmu.edu

Abstract

Adversary emulation is commonly used to test cyber defense performance against known threats to organizations. However, designing attack strategies is an expensive and unreliable manual process, based on subjective evaluation of the state of a network. In this paper, we propose the design of adversarial human-like cognitive models that are dynamic, adaptable, and have the ability to learn from experience. A cognitive model is built according to the theoretical principles of Instance-Based Learning Theory (IBLT) of experiential choice in dynamic tasks. In a simulation experiment, we compared the predictions of an IBL attacker with a carefully designed efficient but deterministic attacker attempting to access an operational server in a network. The results suggest that an IBL cognitive model that emulates human behavior can be a more challenging adversary for defenders than the carefully crafted optimal attack strategies. These insights can be used to inform future adversary emulation efforts and cyber defender training.

Keywords: Cyber security, Adversary emulation, Instance-Based Learning Theory, Cognitive models

1. Introduction

Cyber systems have gradually populated all the personal and collective layers of society. From banks to hospitals, from electric grids to industrial facilities, the interconnectivity of systems has created new opportunities for criminals. Cyber security is a domain of great complexity, defined by uncertainty, lack of visibility, extreme speeds, and partial information. In this adversarial context, defenders and attackers confront each other using digital weapons that are beyond the limits of human capabilities for perception and assessment. Defenders need extensive experience to effectively defend against dynamic and distributed attacks.

Cyber wargaming and adversary emulation (i.e., Red teams) are common practices in organizations to train defenders (i.e., Blue teams) and to develop appropriate defense algorithms (Colbert et al., 2020; Ferguson-Walter et al., 2018). However, the design of emulated adversaries can be expensive and time consuming, especially for scaled networks with a large attack surface and rich defense arsenals. Autonomous agents have been developed to mitigate this problem (Applebaum et al., 2016; Shandilya et al., 2022; Theron et al., 2018).

In particular, game theory has served as an important computational aid in automating the generation of cyber defense strategies. However, these strategies often rely on assumptions of static environments and parameters, including perfect availability of information; and perfect rationality of decision makers (attackers and defenders alike). These are not realistic or useful assumptions, if one wants to generate realistic attack or defense algorithms (Abbasi et al., 2015). Although efforts have been made to relax these assumptions (Denning, 2014; Ferguson-Walter et al., 2019), the design of agents that can represent the strategies of attackers or defenders continues to be challenging. Generally, current machine learning techniques for intrusion detection or malware analysis only perform low-level analyst tasks and often result in large errors and false alarms that can confuse the defense team (Gonzalez et al., 2014). These systems often require continuous retraining and extensive fine-tuning, which is time consuming and inconsistent with the growing need for adaptability and responsiveness against novel threats (Apruzzese et al.,

URI: https://hdl.handle.net/10125/102736
978-0-9981331-6-4

2018; Rosenberg et al., 2021).

These limitations, combined with the rapidly evolving capabilities of cyber attackers and the rise of intelligent autonomous and environmentally aware malware (Thanh & Zelinka, 2019; Theron et al., 2018) present cyber security experts and researchers with a key challenge of developing intelligent defense systems that can learn and understand strategies of dynamic attackers and preempt their intrusions (Dhir et al., 2021).

Cognitive models have been used as embedded computational agents to simulate human interactions with software and networks (Gonzalez et al., 2014; Mitsopoulos et al., 2021; Veksler et al., 2020). Previous work has focused independently on understanding defense behaviors and developing cognitive models of blue agents (Du et al., 2022; Dutt et al., 2011), or the attack preferences and biases of the attacker (Cranford et al., 2020). However, the attacker and the defender are influenced by each other in adversarial cyber scenarios (West & Lebiere, 2001), and such dynamics between attackers and defenders can make defenders more vulnerable to such adversarial actions compared to even random attackers (Moisan & Gonzalez, 2017).

Cognitive agents based on the well-established Instance-Based Learning Theory (IBLT) (Gonzalez et al., 2003), a cognitive theory of decision based on experience, have been used to model the cognitive processes of cyber defenders. Dutt et al. (2011) proposed an IBL model that accurately represents the cyber situation awareness of a human analyst, making concrete predictions of the recognition and comprehension processes of a security expert in a cyber attack. A more recent model from Du et al. (2022) uses a cyber security scenario in which the IBL agent learned to defeat the most aggressive deterministic attack strategy, called *Beeline*.

Humans were also found to handle random attacks better than adaptive attackers in a simple, abstract game (Moisan & Gonzalez, 2017). This suggests that commonly used random-based security algorithms may be less effective than human-inspired adaptive defense strategies. In a phishing experiment, Rajivan and Gonzalez (2018) have found that individual creativity is a predictor of an adversary's ability to evade detection. Cognitive biases and emotions are also believed to affect cyber behaviors, decision making, and strategies of attackers (Ferguson-Walter et al., 2021; Johnson et al., 2021). We hypothesize that cognitive models of *human* adversaries can be more useful in training cyber defenders than deterministic attacker strategies.

In this paper, we propose a cognitive model of a dynamic red agent using the theoretical principles of IBLT. In a cyber security simulation experiment, we compared the performance of the IBL_{Red} agent with that of a deterministic, highly accurate and targeted $Beeline_{Red}$, in attacking a network defended by a dynamic IBL_{Blue} agent. We first developed the IBL_{Red} attacker and trained it in a cyber security scenario against a dormant defender. We then tested this attacker and the best deterministic one ($Beeline_{Red}$) against a dynamic defender developed previously (IBL_{Blue}).

2. Instance-Based Learning Theory

IBLT is a cognitive theory of decision making. It is based on the idea that decisions are made by recognizing similar past experiences, integrating them into the generation of the expected utility of decision alternatives, and selecting the alternative with the maximum expected utility. The development of cognitive models for cyber defense is based on a large body of work on cognitive models of cyber defenders, cyber attackers, and end users in a cyber security context (e.g., Gonzalez et al., 2020).

Although both the process and the mechanisms of IBLT have been published, we repeat the mathematical formulations of the theory here for completeness. The central element of IBLT is the "instance". It represents a unit of memory resulting from the evaluation of potential choice alternatives. Each decision is stored in an instance, structured with three elements that are built over time: a situation state s which is composed of a set of features f; a decision or action a taken corresponding to an alternative in state s; and an expected utility or experienced outcome x of the action taken in a state. Concretely, for an IBL agent, an option $k = (s, a)$ is defined by action a in state s. At time t, assume that n_{kt} different instances (k_i, x_{ik_it}) for $i = 1, ..., n_{kt}$, associated with k. Each instance i in memory has an *Activation* value, which represents the ease of retrieving this information from memory (Anderson & Lebiere, 1998). Here, we consider a simplified version of the Activation equation which only captures recency, frequency, and noise in memory:

$$\Lambda_{ik_it} = \ln\left(\sum_{t' \in T_{ik_it}} (t - t')^{-d}\right) + \sigma \ln \frac{1-\xi_{ik_it}}{\xi_{ik_it}}, \quad (1)$$

where d and σ are the decay and noise parameters, respectively, and $T_{ik_it} \subset \{0, ..., t-1\}$ is the set of previous timestamps in which the instance i was observed. The rightmost term represents a noise for capturing individual variation in activation, and ξ_{ik_it} is a random number drawn from a uniform distribution

$U(0,1)$ at each step and for each instance and option.

Activation of an instance i is used to determine the probability of retrieving an instance from memory. The probability of an instance i is defined by a soft-max function:

$$P_{ik_it} = \frac{e^{\Lambda_{ik_it}/\tau}}{\sum_{j=1}^{n_{kt}} e^{\Lambda_{jk_jt}/\tau}}, \quad (2)$$

where τ is the Boltzmann constant (i.e., the "temperature") in the Boltzmann distribution. For simplicity, τ is often defined as a function of the same σ used in the activation equation $\tau = \sigma\sqrt{2}$.

The expected utility of option k is calculated based on *Blending* as specified in discrete choice tasks (Gonzalez & Dutt, 2011):

$$V_{kt} = \sum_{i=1}^{n_{kt}} P_{ik_it} x_{ik_it}. \quad (3)$$

The choice rule is to select the option that corresponds to the maximum blended value. When the agent receives results that are delayed, the agent updates the expected utilities using a credit assignment mechanism (Nguyen et al., 2021).

3. Cyber Security Scenario

Testing the attacker and defender agents requires a simulation or training platform that encapsulates cyber elements in an integrated environment. On such a platform, defense agents can confront attack agents in cyber scenarios and network simulations. Here, we use the CybORG AI gym (Baillie et al., 2020; Brockman et al., 2016; Standen et al., 2021b) with adversarial cyber operations scenarios to allow users to train agents in a simple but realistic environment. We adapt the CAGE cyber defense scenario (Standen et al., 2021a) to perform experimental simulations using IBL agents as cyber defenders and cyber attackers. This framework was also presented in (Du et al., 2022; Prébot et al., 2022) and in the following we outline its main structural elements and the particularities of the cyber defense scenario.

The attacker (hereafter the Red agent) interacts with the environment through high-level actions that aim to progress and impact the network; the defender (hereafter the Blue agent) aims to stop the progression of the attacker and remove it from the network.

Fig. 1 illustrates the topology of the network chosen for this scenario. The network is divided into three subnets: subnet 1 consists of user hosts that are not critical, subnet 2 consists of enterprise servers designed to support the user activities on Subnet 1, and subnet 3 contains the critical operational server and operational hosts.

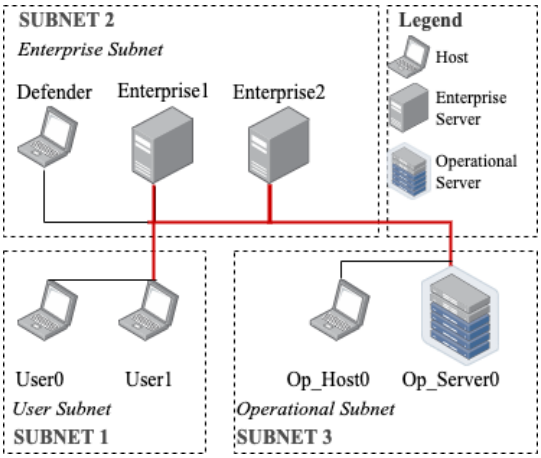

Figure 1. Adaptation of the Cage Challenge Network

Fig.2 summarizes the phases of a targeted attack led by the Red agent (red arrows) and countermeasures for the Blue agent to stop it (blue arrows). The Red agent starts by searching for hosts on the network with *DiscoverRemoteSystems*. To identify vulnerabilities in a target host, the next step is to *DiscoverNetworkServices*. A successful *ExploitRemoteService* on target can obtain *User* level access for the Red agent, which can be escalated to a more privileged *Root* level by *PrivilegeEscalate*. The Blue agent can *Remove* its adversary at the *User* level and use *Restore* if the Red agent has escalated. It can also *Analyse* the activities for additional information or passively *Monitor* the network.

3.1. Red Agent

We used two types of red agents: (1) a deterministic agent, $Beeline_{Red}$, and (2) a dynamic agent, IBL_{Red}. Both red agents start at the host *User0* as their network entry point.

The $Beeline_{Red}$ was proposed in the Cage Challenge scenario (Standen et al., 2021b), and it assumes that the attacker has prior knowledge of the network layout and moves directly to the operational server following the red path ($User0 \rightarrow User1 \rightarrow Enterprise1 \rightarrow Enterprise2 \rightarrow Op\_Server0$) (see Fig. 1) in a predictive and deterministic way.

In contrast, IBL_{Red}, a novel contribution of this research, is a dynamic cognitive agent that learns from experience. This is a cognitive model built according to IBLT to represent human-like memory-based decisions that can adapt its actions dynamically, according to the

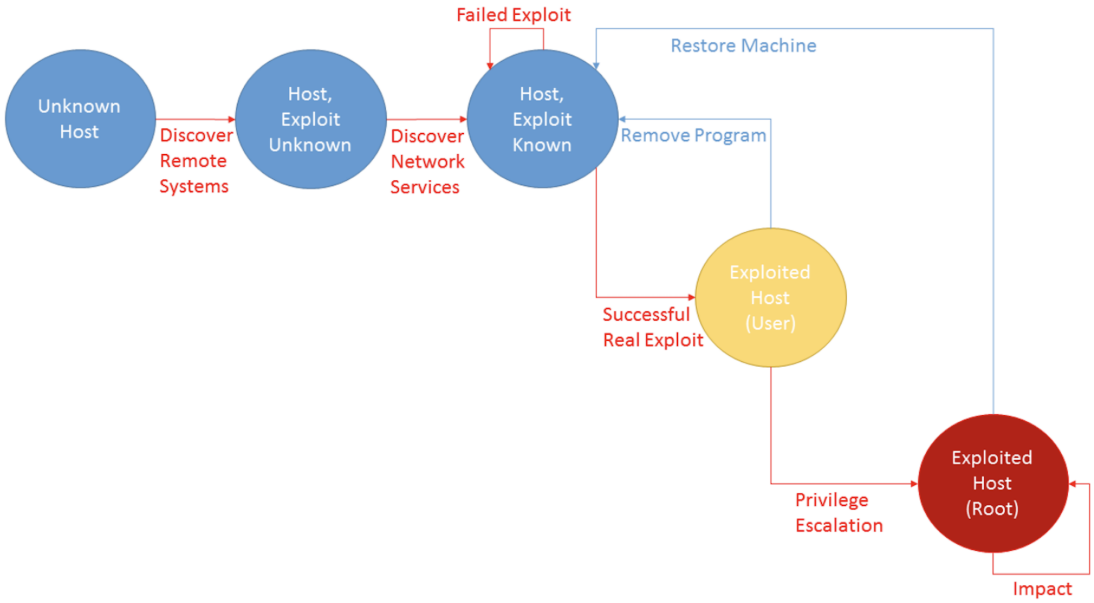

Figure 2. Effect of actions on the host state (diagram from Standen et al., 2021a)

conditions of the environment and the actions of the blue agent. The instances represent each decision made and are structured with the following three elements.

State, s: The state of the instances of the IBL_{Red} agent is composed of features, f, constructed using the concept of Attack Models and Attack Graphs introduced by Sheyner et al. (2002) to model the security vulnerabilities of a network and their exploitation from the perspective of an attacker. Specifically, contextual characteristics include the success status of the previous action of the IBL_{Red} agent and the resources it occupied. A slot is dedicated to each type of resource in various states as shown in Fig. 2. Specifically, a subnet can be newly *Detected* or already *Scanned*, while hosts are classified as *Detected*, *Scanned*, *Exploited (User)*, *Exploited (Root)*, *Impacted*.

The starting status denotes when the IBL_{Red} agent has just successfully established its foothold on the network on *User0*. At that point, only the *User* subnet is detected in addition to its entry point *User0*, while the rest of the slots are empty. The most successful final state for the IBL_{Red} agent is where all hosts and servers are exploited at the *Root* level and when the critical *Op_Server0* is impacted.

Action Space, a: The action space for the IBL_{Red} agent is dynamically constructed at each step based on the status of each host in the network. Each action consists of a target host and an applicable command. As shown in Fig. 2, IBL_{Red} can choose to *Discover** more hosts in the network, or advance the attack status of known hosts.

Utility, z: A reward is calculated at each step, based on the attack status, as shown in Table 1. Higher rewards are assigned when the IBL_{Red} agent is able to access more significant systems. Only root access to the systems and successful impact on the operational server are rewarded. The IBL_{Red} agent receives a reward of 0 for any other action.

Event or Action	Reward
Administrator access on a Host	0.1
Administrator access on a Server	1
Successfully *Impact Op_Server0*	10

Table 1. Events and actions costs

3.2. Blue Agent

We used two types of blue agents: (1) a deterministic agent $Sleepy_{Blue}$ and (2) a dynamic agent IBL_{Blue}. $Sleepy_{Blue}$ passively *Monitors* the network and does not attempt to stop the Red agent. The IBL_{Blue}, has been proposed and tested recently by Du et al. (2022). This agent, also built according to IBLT, learns from experience and adapts its actions dynamically, according to the conditions of the environment and the actions of the red agent. Experiments demonstrate that IBL_{Blue} provided with delayed feedback learn

to exploit the deterministic nature of $Beeline_{Red}$ and achieve near-zero loss.

At each step, based on the observed state of the network and the consequences of the attacker's previous actions, the blue agent selects a host or server to act on and one of four possible actions: *Analyze* is used to collect information about the level of compromise of the selected host; *Remove* is used to remove a suspected malicious agent from the host or server; if the malicious agent cannot be removed, the blue agent can *Restore* a host or server to a previous stable state; and *Monitor* to just continue observing the system, which has essentially no effect on the network state.

4. Hypotheses and Simulation Methods

Hypothesis 1. The first hypothesis tests the behavior of two red agents: the $Beeline_{Red}$ agent and the IBL_{Red} agent, against the $Sleepy_{Blue}$ agent.

We expect that: (a) our newly proposed IBL_{Red} agent will learn from experience, and will achieve a similar level of performance as the $Beeline_{Red}$, the best known strategy in the Adapted Cage Challenge scenario, after learning; (b) the $Beeline_{Red}$ agent will consistently receive the highest reward and maximum impact on the operational server, since $Beeline_{Red}$ represents a deterministic but highly effective attacker; and (c) the IBL_{Red} agent will initially perform poorly, since it can only learn from experience, but would learn to take advantage of such an ineffective defender with practice.

Hypothesis 2. The second hypothesis tests the behavior of two red agents: the $Beeline_{Red}$ agent and the $IBL_{Red}^{Trained}$ agent, against the IBL_{Blue} agent (Du et al., 2022). $IBL_{Red}^{Trained}$ is IBL_{Red} pre-trained against $Sleepy_{Blue}$ for 2000 episodes. This type of agent is designed to simulate the advanced stealthy threat actors in the real world who gain unauthorized access to a computer network and remain undetected for an extended period when the cyber defender is still "sleeping".

We expect $Beeline_{Red}$ to initially achieve higher reward and longer impact duration than $IBL_{Red}^{Trained}$ agent. However, the determinism and static nature of $Beeline_{Red}$ will be exploited by the learning IBL_{Blue} agent, resulting in worse attack performance of $Beeline_{Red}$ than $IBL_{Red}^{Trained}$.

Methods. Each of the two hypotheses were tested in separate simulation experiments. We ran 40 IBL runs, each with 2000 episodes. The duration of the episode was set to 25 steps to ensure that the Blue agent could fully observe the attack strategies. All IBL models were run with default decay $d = 0.5$ and noise $\sigma = 0.25$. This means that the results presented here are all a priori predictions of how a human defender IBL_{Blue} and a human attacker IBL_{Red} are expected to behave under these conditions.

The performance of the Red agent was evaluated for each episode, using the following metrics: **(1) Reward:** the reward received during the execution of the scenario; **(2) Impact duration:** the average number of steps per episode that the Red agent successfully impacts the operational server; **(3) Progress:** the average number of steps per episode that the Red agent took to penetrate the *Enterprise subnet* and *Operational subnet*; and **(4) Action frequency:** the average proportion of command usage at each step in an episode.

The performance of the Blue agent was also evaluated in terms of: **(1) Action frequency:** proportions of command usages at each step; and **(2) Number of options:** the average number of defense choices available to the blue agent. This represents the decision space left to the defender after each action of the Red agent. Options are combinations of blue commands and hosts.

5. Results

5.1. Red Agent Performance

Attacker Reward. Fig. 3-Left panel shows the test of Hypothesis 1 and Fig. 3-Right panel the test of Hypothesis 2 in terms of the reward obtained by the Red agents. To test the observations, we ran a one-way between subjects ANOVA using *attacker type* as the main factor and *attacker reward* as the dependent variable, aggregating for the first 500 and the last 500 episodes.

As expected in Hypothesis 1, the IBL_{Red} agents ($M = 59.09, SD = 43.78$) performed significantly worse than the $Beeline_{Red}$ agents ($M = 112.8$, $SD = 0$) when faced with $Sleepy_{Blue}$ in the first 500 episodes [$F(1, 39998) = 30105, p < .001, \eta^2 = 0.43$]. In contrast, the IBL_{Red} agents ($M = 104.64$, $SD = 38.68$) are able to reach comparable average performance as the $Beeline_{Red}$ agents ($M = 112.8$, $SD = 0$) in the last 500 episodes By the end of the 2000th episode, 55% of the IBL_{Red} agents received a higher reward than $Beeline_{Red}$, which requires them to quickly penetrate the network to Impact *Op_Server0*, and at the same time fully exploit the remaining valuable systems when the opportunity arises. Most importantly, the IBL_{Red} agent learned such a complex and efficient strategy purely from experience according

to IBLT (Gonzalez et al., 2003) and without any explicit encoding of any strategy.

Also, as expected in Hypothesis 2, given that the IBL_{Blue} agent also starts naively learning from experience, the $Beeline_{Red}$ agents performed better ($M = 112.8$, $SD = 30.39$) than the $IBL^{Trained}_{Red}$ agents ($M = 80.34$, $SD = 45.02$) in the first 500 episodes [$F(1, 39998) = 20579, p < .001, \eta^2 = 0.34$]. However, the $IBL^{Trained}_{Red}$ agents posed a more persistent threat than the $Beeline_{Red}$ agents, while the performance of the $Beeline_{Red}$ agents deteriorates rapidly. The $Beeline_{Red}$ agents ($M = 54.60$, $SD = 34.31$) performed significantly worse than the $IBL^{Trained}_{Red}$ agents ($M = 5.15$, $SD = 14.68$) in the last 500 episodes [$F(1, 39998) = 31185, p < .001, \eta^2 = 0.44$].

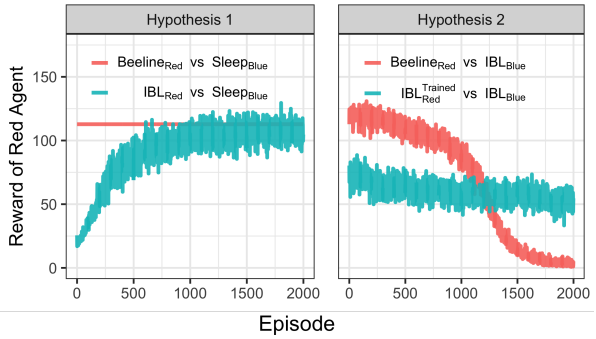

Figure 3. Reward

Impact duration. The main goal of the attacker is to maintain constant *Impact* over *Op_Server0*. Fig. 4-Left panel shows the test of Hypothesis 1 and 4-Right panel shows the test of Hypothesis 2, in terms of the number of successive impacts performed by the red agent on the *Op_Server*

As expected in Hypothesis 1, the IBL_{Red} agents are capable of achieving a comparable impact on the network ($M = 9.0$, $SD = 3.37$) as the $Beeline_{Red}$ agent ($M = 6.4$, $SD = 1$) when faced with $Sleepy_{Blue}$ in the last 500 episodes [$F(1, 39998) = 901.4, p < .001, \eta^2 = 0.31$].

Also, as expected in Hypothesis 2, $IBL^{Trained}_{Red}$ achieves shorter impact duration ($M = 5.08$, $SD = 3.77$) than $Beeline_{Red}$ ($M = 8.93$, $SD = 2.34$) in the first 500 episodes [$F(1, 39998) = 17240, p < .001, \eta^2 = 0.95$]. This relative disadvantage reversed in last 500 episodes, where the $IBL^{Trained}_{Red}$ had a higher impact duration ($M = 3.25$, $SD = 2.82$) than the $Beeline_{Red}$: ($M = 0.18$, $SD = 1.11$) [$F(1, 39998) = 14699, p < .001, \eta^2 = 0.94$].

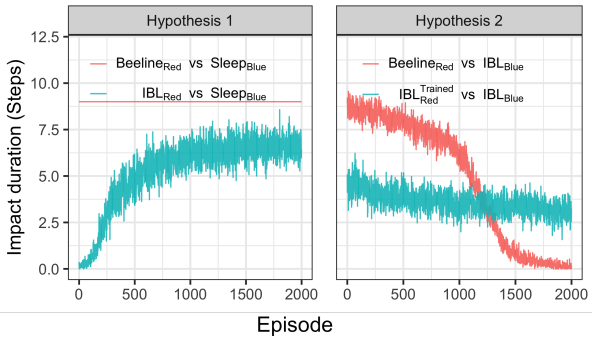

Figure 4. Impact duration: The number of successive impacts performed by the red agent on Op_Server

5.2. Further Exploration of Red Agents' Behavior

To further explore the behavior of red agents with respect to Hypothesis 2, we analyzed the number of steps that the attacker takes to reach a subnet (Enterprise and Operational). Considering the layered network structure shown in Fig. 1, the progress of the Red agents can be marked by two milestones: penetration of the *Enterprise* subnet and the *Operational* subnet. As expected in Hypothesis 2, we can observe the increasingly delayed and impeded forward progress of $Beeline_{Red}$ from Fig. 5.

$Beeline_{Red}$ takes on average 15 more steps to enter the *Enterprise* subnet than initially, and longer for the *Operational* one too. However, $IBL^{Trained}_{Red}$ shows relatively stable performance over the course of the 2000 episodes and as IBL_{Blue} gains experience. $IBL^{Trained}_{Red}$ takes longer time to penetrate the Enterprise subnet ($M = 4.77$, $SD = 0.33$) and the Operational subnet ($M = 15.67$, $D = 0.61$) than $Beeline_{Red}$ (Enterprise: ($M = 4.59$, $SD = 0.33$), Operational: ($M = 12.04$, $D = 0.38$)) in the first 500 episodes [Enterprise: $F(1, 39998) = 66.02, p < .001, \eta^2 = 0.06$] [Operational: $F(1, 39998) = 12489, p < .001, \eta^2 = 0..93$]. This relative disadvantage reversed in last 500 episodes, where the $IBL^{Trained}_{Red}$ propagated faster into Enterprise subnet ($M = 6.12$, $SD = 0.40$) and the Operational subnet ($M = 16.90$, $D = 0.52$) than the $Beeline_{Red}$: Enterprise: ($M = 19.20$, $SD = 0.99$) [$F(1, 39998) = 74265, p < .001, \eta^2 = 0.99$], Operational: ($M = 21.27$, $D = 0.97$), [$F(1, 39998) = 7940, p < .001, \eta^2 = 0.89$].

Fig. 6 compares the average distribution of the use of attack commands at each step of the first 500 episodes (Left panels) in Stage 2 versus the last 200 episodes

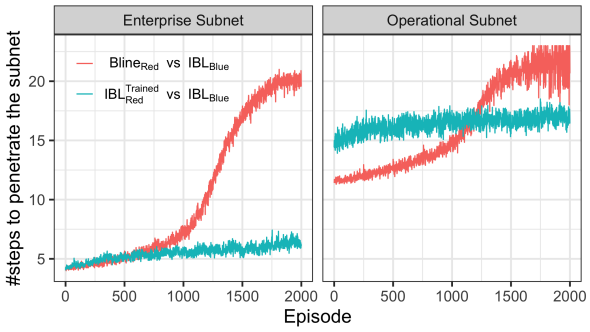

Figure 5. Hypothesis 2 - Progress: The number of steps it takes for the attacker to reach subnet Enterprise and Op

(Right panels). $IBL_{Red}^{Trained}$ and $Beeline_{Red}$ present similar proportions of actions at the beginning, with higher *Sleep* proportion for $IBL_{Red}^{Trained}$. The difference becomes much larger in the final episodes. $Beeline_{Red}$ is stuck into a loop of *ExploitRemoteService* and *PrivilegeEscalate*, while $IBL_{Red}^{Trained}$ maintained a consistent distribution. This comparison constitutes further evidence of the inefficacy of deterministic heuristic strategies. The disappearance of *Discover** actions and *Impact* actions can help explain the reason for the rapid drop in reward within the episodes.

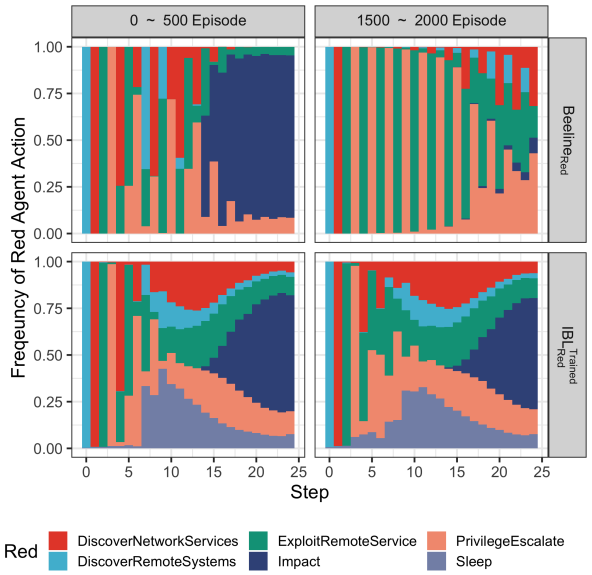

Figure 6. Average frequency of attacker command usage at each step in an episode

5.3. Exploration of IBL_{Blue} agent's Behavior

The Blue agent performance in terms of *Reward* and *Impact duration* has been evaluated in (Du et al., 2022). We are not going to repeat those results here, and they are essentially reversed results for the IBL_{Blue} and $Beeline_{Red}$ interactions we presented above. Instead, we will focus this section on the exploration of Hypothesis 2 from the defender's side.

As presented in Fig. 7, the dynamics of the use of defensive commands by the agent IBL_{Blue} shows a difference when confronting the agent $Beeline_{Red}$ in contrast to the agent IBL_{Red}. IBL_{Blue} agents faced with a $Beeline_{Red}$ attacker are able to minimize the proportion of costly *Restore* action and stop the attacker with *Remove* in an earlier state of the cyber-kill chain. Those fighting with $IBL_{Red}^{Trained}$ failed to derive a defense strategy better than a random defense.

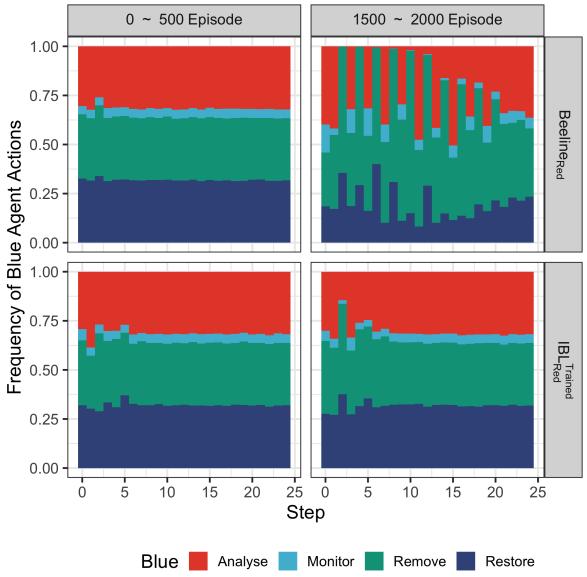

Figure 7. Average frequency of defender command usage at each step in an episode

Size of the Option Space: Psychology and behavioral research suggests that too many choices can overload decision makers (Reed et al., 2012; Schwartz & Ward, 2004). Defenders face this challenge of information overload. Fig. 8 analyzes the number of options available to the IBL_{Blue} agent during the 25 steps of the episodes. As shown in the left panel, when IBL_{Blue} fought against $Beeline_{Red}$, it was able to reduce the option space in the final 500 episodes compared to the first 500 episodes.

In contrast, Fig. 8-Right panel shows that the option

space stays about the same size from the first 500 episodes to the last 500 episodes. That is, the IBL_{Blue} agent was unable to simplify the option space with experience against the $IBL_{Red}^{Trained}$ agent by impeding its progress and minimize the number of exploited hosts.

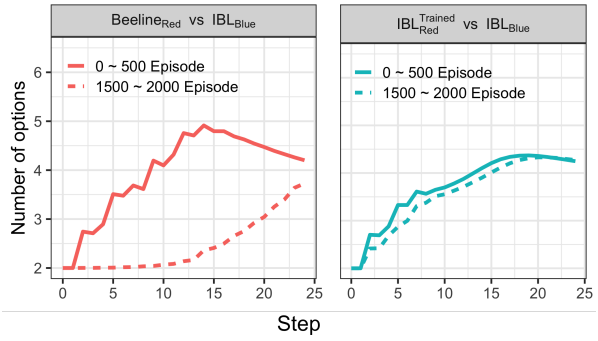

Figure 8. Number of choice options for defender

6. Discussion

Adversary emulation strategies can be used to train cyber defense teams, develop intelligent systems for cyber defense, and test cyber defense capabilities. However, the process of developing effective adversary emulations can be expensive and their evaluation is often subjective (Russo et al., 2019; Yoo et al., 2020). A first contribution of this paper is to demonstrate that it is possible to evaluate cyber defense intelligent systems by generating a "cyber-war between bots"; where automated cyber defenders can be paired with automated cyber attackers to determine their performance and strategy.

Second, we demonstrate that cognitive models, aimed at emulating human cognitive processes of decision making, can be instrumental in generating defense automation capabilities (Gonzalez et al., 2014). Concretely, we demonstrate that cognitive models that emulate human adversaries can be better test cases for cyber defense teams and for technological capabilities. We present a cognitive model of an attacker based on IBLT (Gonzalez et al., 2003), IBL_{Red}. IBL_{Red} is first trained against a static and inactive defender, $Sleep_{Blue}$. The main feature of IBL_{Red} is that it can learn from interactive feedback on the task, and we showed that it can reach the same level of effectiveness as the best adversarial strategy in this scenario, $Beeline_{Red}$. This result suggests that an IBL cognitive model can be an effective dynamic and adaptive emulator of attack strategies. Importantly, the IBL attacker can adapt and learn according to the dynamics of the cyber defense environment.

A third contribution of this paper is to demonstrate the performance of an IBL model of the defender IBL_{Blue} when paired with two different types of emulated attackers: the optimal attack strategy $Beeline_{Red}$, and a human-like attacker $IBL_{Red}^{Trained}$. This IBL_{Blue} defender can learn the optimal strategy again, but a human-like attack strategy $IBL_{Red}^{Trained}$ is more difficult for the IBL_{Blue} defender to learn than an optimal but stable optimal attack strategy. The explanation is that using a cognitive model to emulate attackers is more effective than using deterministic strategies. Cognitive models are dynamic and adaptive to the defender's actions, while the Beeline strategy is static and consistent. The IBL_{Blue} agent was able to learn the Beeline strategy and eventually take advantage of it, while it did not effectively hinder the progress of the IBL_{Red} agent.

Our analyses show that it takes significantly more steps over time for the Beeline attacker to reach the Enterprise subnet and ultimately more steps to reach the Operational server. The IBL_{Blue} learns over time to prevent these actions from this Beeline strategy. However, it is significantly more difficult to prevent IBL_{Red} from reaching the Enterprise and the Operational servers. We further verify that there is important learning that occurs from the first to the last episodes in terms of the actions taken by the attacker. For example, the number of impact actions is significantly reduced from the first to the last 500 episodes when the IBL_{Blue} agent confronts the Beeline strategy, but the reduction in impact actions is minimal when the IBL_{Blue} agent confronts the human-like IBL_{Red} agent.

Exploring the actions taken by the IBL_{Blue} agent suggests that the agent learns to decrease restore actions when confronted with the agent $Beeline_{Red}$, while maintaining a more consistent distribution of actions when confronted with the agent IBL_{Red}. When analyzing the options with which the IBL_{Blue} agent is confronted at each particular time, we observed an interesting effect: the IBL_{Blue} agent learned to reduce its decision option space against the $Beeline_{Red}$, while the option space of the IBL_{Blue} agent against the IBL_{Red} did not decrease substantially.

6.1. Conclusion and Limitations

In conclusion, we provide important steps towards establishing emulated adversaries that can be effective for training cyber defenders and supporting the development of autonomous cyber defenders. We demonstrate that it is possible to use cognitive models that emulate human-like strategies to produce

adversaries that are adaptive to the actions of defenders. These models can ultimately be more effective in learning cyber defense strategies than static and deterministic adversaries. However, demonstrating the benefits of the use of cognitive models in real-world cyber security environments remains a research challenge. Extending the scenario to the size of real-world networks can exponentially expand the state space in the cognitive model, and research on partially observable states for the defender will be required to account for imperfect network monitoring infrastructures. Future work will aim at verifying the predictions of the effectiveness of human defenders when confronted with these two types of attack strategies, and to improving the game model to be more representative of real-world environments.

7. Acknowledgement

This research was sponsored by the Army Research Office and accomplished under Australia-US MURI Grant Number W911NF-20-S-000 and by the Army Research Laboratory under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA).

References

Abbasi, Y. D., Short, M., Sinha, A., Sintov, N., Zhang, C., & Tambe, M. (2015). Human adversaries in opportunistic crime security games: Evaluating competing bounded rationality models. *Proceedings of the third annual conference on advances in cognitive systems ACS*, 2.

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought* (J. R. Anderson & C. J. Lebiere, Eds.). Psychology Press. https://doi.org/10.4324/9781315805696

Applebaum, A., Miller, D., Strom, B., Korban, C., & Wolf, R. (2016). Intelligent, automated red team emulation. *Proceedings of the 32nd Annual Conference on Computer Security Applications*, 363–373.

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. *10th international conference on cyber Conflict (CyCon)*, 371–390.

Baillie, C., Standen, M., Schwartz, J., Docking, M., Bowman, D., & Kim, J. (2020). Cyborg: An autonomous cyber operations research gym. *arXiv:2002.10667*.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym. *ArXiv*, *abs/1606.01540*.

Colbert, E. J., Kott, A., & Knachel, L. P. (2020). The game-theoretic model and experimental investigation of cyber wargaming. *The Journal of Defense Modeling and Simulation*, *17*(1), 21–38.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., & Lebiere, C. (2020). Toward personalized deceptive signaling for cyber defense using cognitive models. *Topics in Cognitive Science*, *12*(3), 992–1011.

Denning, D. E. (2014). Framework and principles for active cyber defense. *Computers & Security*, *40*, 108–113.

Dhir, N., Hoeltgebaum, H., Adams, N., Briers, M., Burke, A., & Jones, P. (2021). Prospective artificial intelligence approaches for active cyber defence.

Du, Y., Prébot, B., Xi, X., & Gonzalez, C. (2022). Towards autonomous cyber defense: Predictions from a cognitive model. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Dutt, V., Ahn, Y.-S., & Gonzalez, C. (2011). Cyber situation awareness: Modeling the security analyst in a cyber-attack scenario through instance-based learning. In Y. Li (Ed.), *Data and applications security and privacy xxv* (pp. 280–292). Springer Berlin Heidelberg.

Ferguson-Walter, K., Fugate, S., Mauger, J., & Major, M. (2019). Game theory for adaptive defensive cyber deception. *Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security*, 1–8.

Ferguson-Walter, K., Gutzwiller, R., Scott, D., & Johnson, C. (2021). Oppositional human factors in cybersecurity: A preliminary analysis of affective states. https://doi.org/10.1109/ASEW52652.2021.00040

Ferguson-Walter, K., Shade, T., Rogers, A., Trumbo, M. C. S., Nauer, K. S., Divis, K. M., Jones, A., Combs, A., & Abbott, R. G. (2018). *The tularosa study: An experimental design and implementation to quantify the effectiveness of cyber deception.* (tech. rep.). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Gonzalez, C., Aggarwal, P., Lebiere, C., & Cranford, E. (2020). Design of dynamic and personalized deception: A research framework and new insights. *Proceedings of the 53rd Hawaii*

International Conference on system sciences, 1825–1834.

Gonzalez, C., Ben-Asher, N., Oltramari, A., & Lebiere, C. (2014). Cognition and technology. In A. Kott, C. Wang, & R. F. Erbacher (Eds.), *Cyber defense and situational awareness* (pp. 93–117). Springer International Publishing. https://doi.org/10.1007/978-3-319-11391-3_6

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological review*, *118*(4), 523.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cogn. Sci.*, *27*, 591–635.

Johnson, C. K., Gutzwiller, R. S., Gervais, J., & Ferguson-Walter, K. J. (2021). Decision-making biases and cyber attackers. *2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, 140–144.

Mitsopoulos, K., Somers, S., Schooler, J., Lebiere, C., Pirolli, P., & Thomson, R. (2021). Toward a psychology of deep reinforcement learning agents using a cognitive architecture. *Topics in Cognitive Science*.

Moisan, F., & Gonzalez, C. (2017). Security under uncertainty: Adaptive attackers are more challenging to human defenders than random attackers. *Frontiers in psychology*, *8*, 982.

Nguyen, T. N., McDonald, C., & Gonzalez, C. (2021). *Credit assignment: Challenges and opportunities in developing human-like ai agents* (tech. rep.). Carnegie Mellon University.

Prébot, B., Du, Y., Xi, X., & Gonzalez, C. (2022). Cognitive models of dynamic decision in autonomous intelligent cyber defense. *2nd International Conference on Autonomous Intelligent Cyber-defence Agents AICA 2022*.

Rajivan, P., & Gonzalez, C. (2018). Creative persuasion: A study on adversarial behaviors and strategies in phishing attacks. *Frontiers in psychology*, *9*, 135.

Reed, D. D., Kaplan, B. A., & Brewer, A. T. (2012). Discounting the freedom to choose: Implications for the paradox of choice. *Behavioural processes*, *90*(3), 424–427.

Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Computing Surveys (CSUR)*, *54*(5), 1–36.

Russo, L., Binaschi, F., De Angelis, A., Armando, A., Henauer, M., & Rigoni, A. (2019). Cybersecurity exercises: Wargaming and red teaming. *Next Generation CERTs*, *54*, 44.

Schwartz, B., & Ward, A. (2004). Doing better but feeling worse: The paradox of choice. *Positive psychology in practice*, 86–104.

Shandilya, S. K., Upadhyay, S., Kumar, A., & Nagar, A. K. (2022). Ai-assisted computer network operations testbed for nature-inspired cyber security based adaptive defense simulation and analysis. *Future Generation Computer Systems*, *127*, 297–308.

Sheyner, O., Haines, J., Jha, S., Lippmann, R., & Wing, J. M. (2002). Automated generation and analysis of attack graphs. *Proceedings 2002 IEEE Symposium on Security and Privacy*, 273–284.

Standen, M., Lucas, M., Bowman, D., Richer, T. J., Kim, J., & Marriott, D. (2021a). Cage challenge 1. *IJCAI-21 1st International Workshop on Adaptive Cyber Defense.*

Standen, M., Lucas, M., Bowman, D., Richer, T. J., Kim, J., & Marriott, D. (2021b). Cyborg: A gym for the development of autonomous cyber agents.

Thanh, C. T., & Zelinka, I. (2019). A survey on artificial intelligence in malware as next-generation threats. *Mendel*, *25*(2), 27–34.

Theron, P., Kott, A., Drašar, M., Rzadca, K., LeBlanc, B., Pihelgas, M., Mancini, L., & Panico, A. (2018). Towards an active, autonomous and intelligent cyber defense of military systems: The nato aica reference architecture. *2018 International conference on military communications and information systems (ICMCIS)*, 1–9.

Veksler, V. D., Buchler, N., LaFleur, C. G., Yu, M. S., Lebiere, C., & Gonzalez, C. (2020). Cognitive models in cybersecurity: Learning from expert analysts and predicting attacker behavior. *Frontiers in Psychology*, *11*, 1049.

West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Cognitive Systems Research*, *1*(4), 221–239.

Yoo, J. D., Park, E., Lee, G., Ahn, M. K., Kim, D., Seo, S., & Kim, H. K. (2020). Cyber attack and defense emulation agents. *Applied Sciences*, *10*(6), 2140.